Supplemental Figure 1

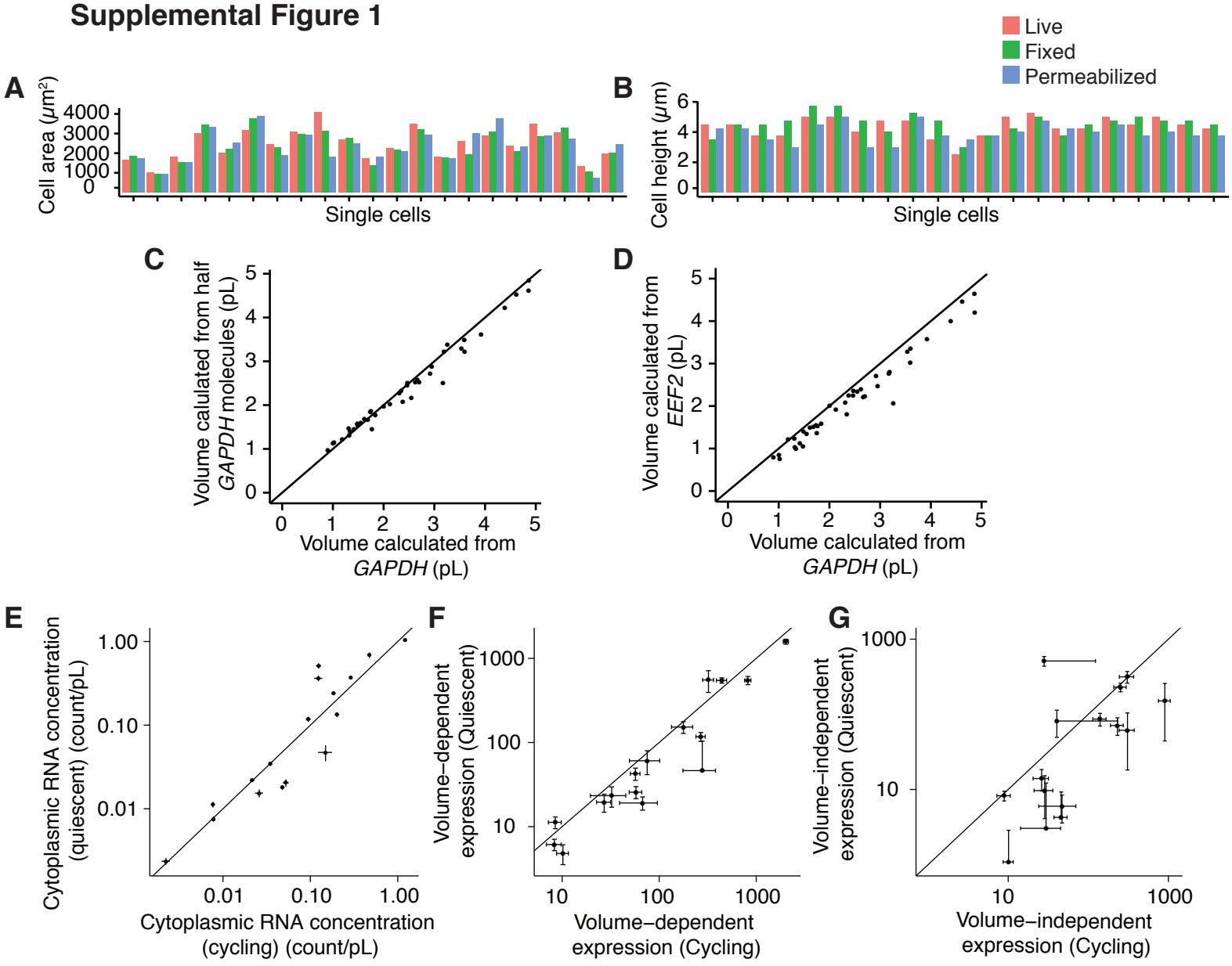

Figure S1, related to Figure 1.

A and B. We monitored primary fibroblast cells on the microscope throughout the process of fixation. We took measurements of the same cells live, after fixing in 4% formaldehyde for 10 minutes, and after permeabilizing in ethanol for 30 minutes.

A. We measured the areas of the cells through brightfield images.

B. We measured the height of the cells by coating the cells with fluorescent beads. These measurements indicate that the cells remain roughly the same size throughout the fixation and permeabilization process.

C. To demonstrate the robustness of the volume calculation algorithm, we calculated volume for the same cells using all the *GAPDH* mRNA spot coordinates as detected by RNA FISH, or using only half of the points, chosen randomly. Both methods result in approximately the same volume, suggesting that the number of points we use is sufficient to calculate the volume accurately.

D. We calculated volume using a different gene, *EEF2*. On average, *EEF2* has an abundance that is less than half that of *GAPDH* (mean EEF2 = 1079 mRNA/cell, mean *GAPDH* = 2673 mRNA/cell). Volume calculated using *EEF2* is systematically lower than that calculated using *GAPDH*, but the values are similar.Black lines in C, D indicate a fit with intercept = 0 and slope = 1.

E. mRNA concentration is similar in cycling and quiescent cells. We calculated the average mRNA concentration for 17 genes in both the cycling and quiescent state in human foreskin fibroblast cells. Each data point represents one gene. Each gene had a minimum of 2 biological replicates, with at least thirty cells per replicate. Line has intercept 0 and slope 1. Error bars represent standard error.

F,G. We compared volume-dependent and -independent abundance for cycling and quiescent cells. Both volume-independent and volume-dependent expression are lower in quiescent cells. All error bars represent confidence intervals of the slope or intercept of the fit, normalized to the scale of the plot. In C and D, we omitted error bars that extended below zero. Each gene had a minimum of two biological replicates, with at least 30 cells per replicate. We omitted highly variable genes with intercept terms less than zero.

Supplemental Figure 2

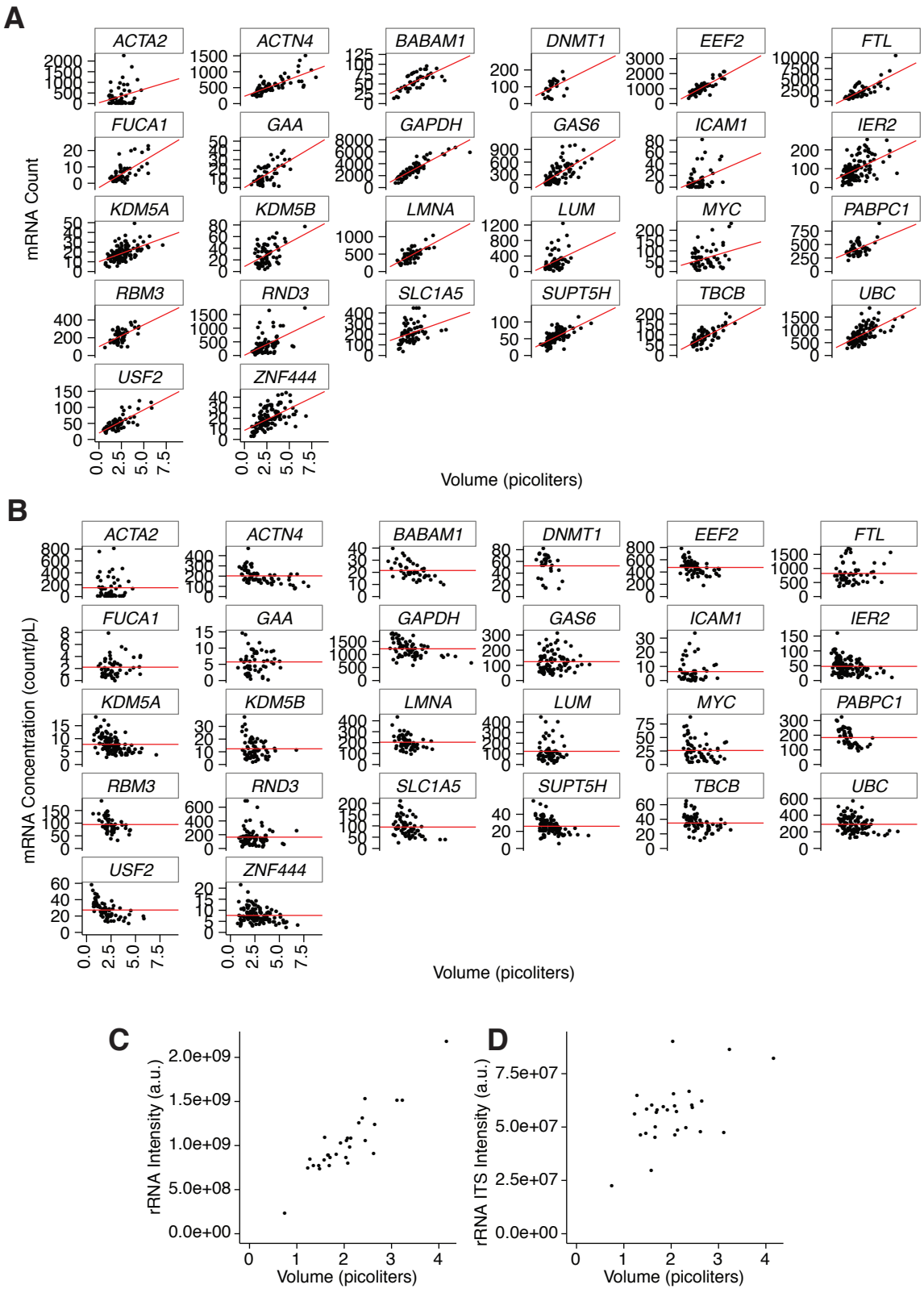

Figure S2, related to Figure 1.

A, B. mRNA count (A) and concentration (B) measurements for all genes in cycling fibroblast cells. Each data point is an individual single cell measurement. In count plots, red line indicates best linear fit to the data. In concentration plots, red line indicates mean mRNA concentration. Each data set is a combination of at least two biological replicates.

C. We measured ribosomal RNA by quantifying total fluorescence intensity in the cytoplasm from an rRNA FISH probe in cycling fibroblast cells.

D. We measured the rRNA ITS (the rRNA "intron") by quantifying total fluorescence intensity in the nucleus from an ITS RNA FISH probe.

rRNA and the rRNA ITS both scale with volume to some degree, suggesting that the production of ribosomal RNA scales with volume. We have shown that mRNA scales with volume, so a similar scaling of rRNA is not inconsistent with the production of protein to scale with volume as well. The data shown for rRNA is one of three biological replicates.

Supplemental Figure 3

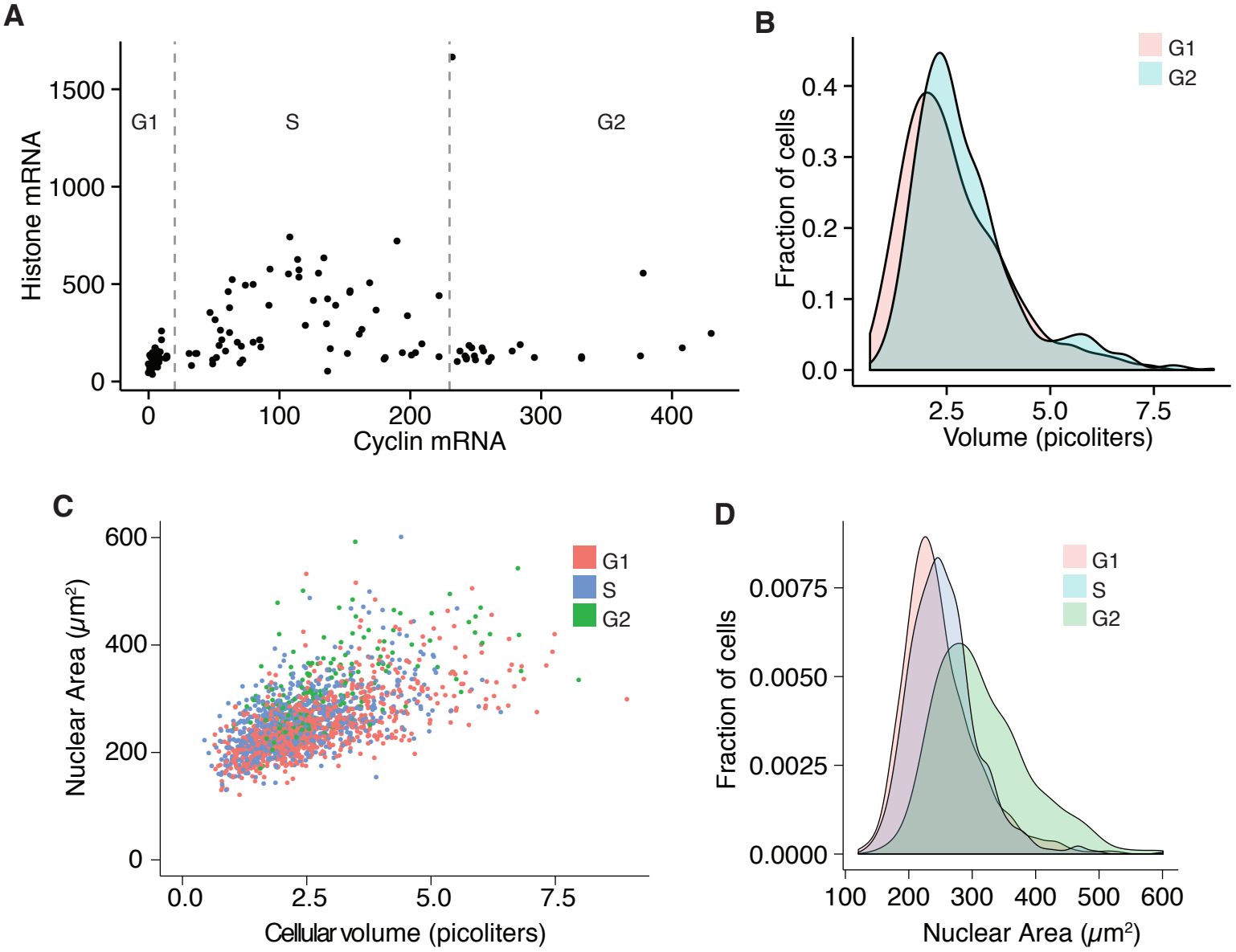

Figure S3, related to Figures 1, 5, 6.
A. We simultaneously measured *CCNA2* and *HIST1H4E* mRNA by RNA FISH to precisely determine cell cycle position. Each data point is a single cell measurement. *CCNA2* is highly expressed in S and G2, but not G1. *HIST1H4E* is highly expressed only in S phase. Cells with low *CCNA2* and *HIST1H4E* are in G1 (cutoff = 20 *CCNA2* mRNA), cells with mid-range *CCNA2* and high *HIST1H4E* are in S, and cells with high *CCNA2* and low *HIST1H4E* are in G2 (cutoff = 230 *CCNA2* mRNA). We determined thresholds for all samples using this method. Data shown are from one of four biological replicates.
B. Volume distributions in G1 and G2. We determine cell cycle position using *CCNA2*. We note that G2 cells are larger than G1 cells, but only 10% larger on average, possibly due to non-linearities in growth over the course of the cell cycle. n = 841 cells in G1, 191 cells in G2.
C. Nuclear area vs. cytoplasmic volume. We measure cytoplasmic volume using our standard method. We measure nuclear area using the DAPI stain. We note that we only measure nuclear area and not volume.
D. Density plot of nuclear area across cell cycle stages. While nuclear area generally scales with cytoplasmic volume, there is considerable spread in the data ($R^2 = 0.358$). n = 1866 cells.

Supplemental Figure 4

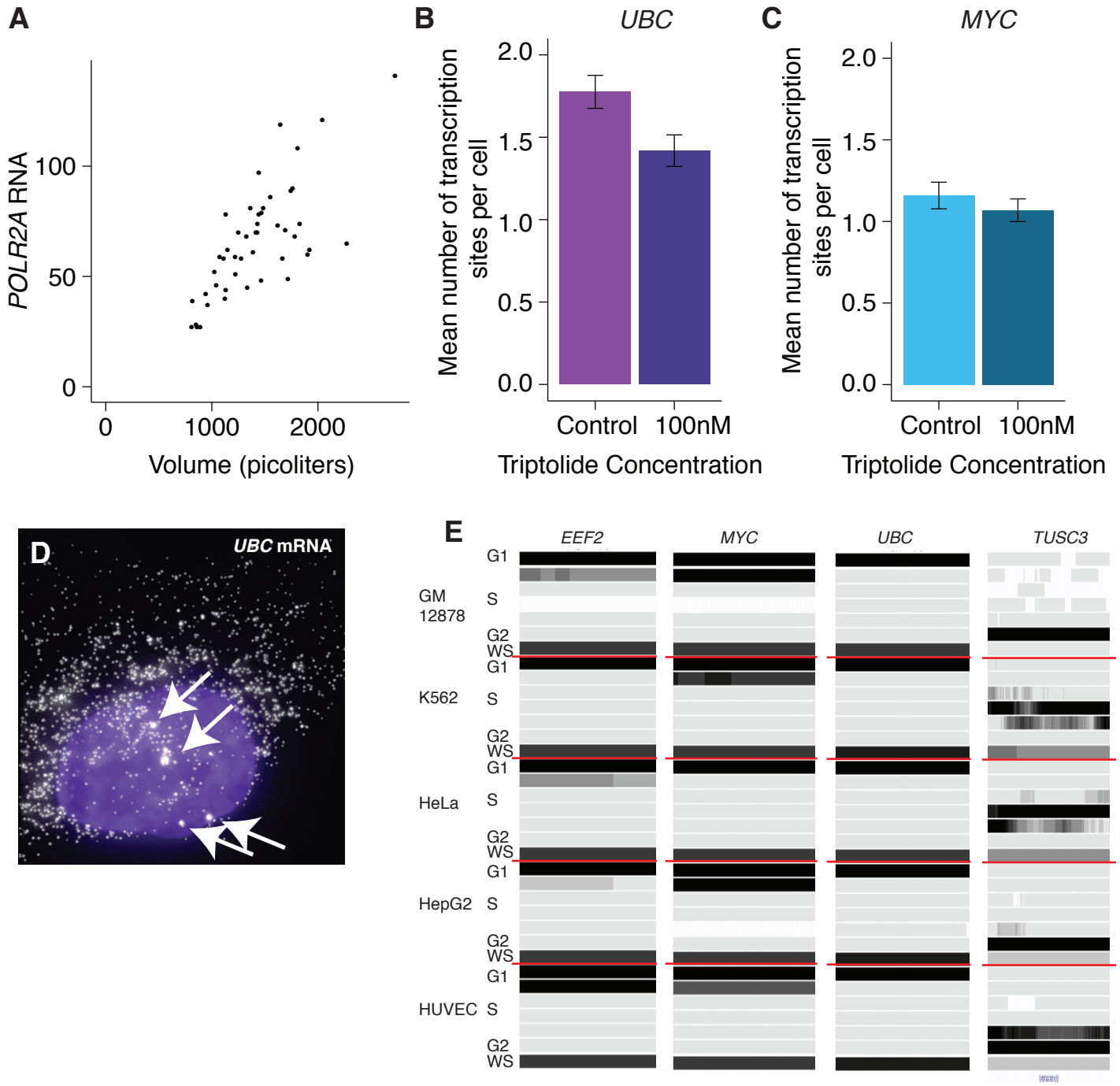

Figure S4, related to Figures 5 and 6.

A. Quantification of RNA Polymerase II mRNA (quantified by RNA FISH) vs. cytoplasmic volume in A549 cells. Data shown is from a single biological replicate.

B,C.We treated primary fibroblast cells with 100nM triptolide (100nM) or simply replaced the media on cells (Control) and fixed in methanol after one hour. For genes *UBC* and *MYC*, we identified active transcription sites through intron/exon colocalization and quantified the number of active transcription sites per cell with and without triptolide. Bars represent mean of cells, error bars are SEM. The data shown here is from two replicates, for a total of 282 cells for *UBC* and 257 cells for *MYC*. For both genes, the change in frequency between conditions is small compared to the change in intensity (Fig. 5).

D. *UBC* mRNA in a CRL2097 cell. RNA FISH probe in white, DAPI stain in purple. White arrows indicate transcription sites. We detect transcription sites through intron/exon colocalization by RNA FISH. This cell is in G2 and has four transcription sites, demonstrating that all gene copies are transcriptionally competent after replication.

E. Tracks from UCSC genome browser displaying UW Repli-Seq data in GM12878 (lymphoblastoid), K562 (chronic myelogenous leukemia), HeLa (cervical cancer), HepG2 (liver carcinoma), and HUVEC (human umbilical vein endothelial) cell lines. The track displays data for different points in the cell cycle: G1, S1 (early S phase), S2 (middle-early S phase), S3 (middle-late S phase), S4 (late S phase), and G2. WS represents a wavelet-smoothed transform of the six other tracks. This data was generated by sequencing newly-replicated DNA in each point in the cell cycle. Darkness of track corresponds to read density. Each track shown corresponds to entire length of each gene. Data is shown for early replicating genes *EEF2*, *MYC*, *UBC*, and a late replicating gene, *TUSC3*.

Supplemental Figure 5

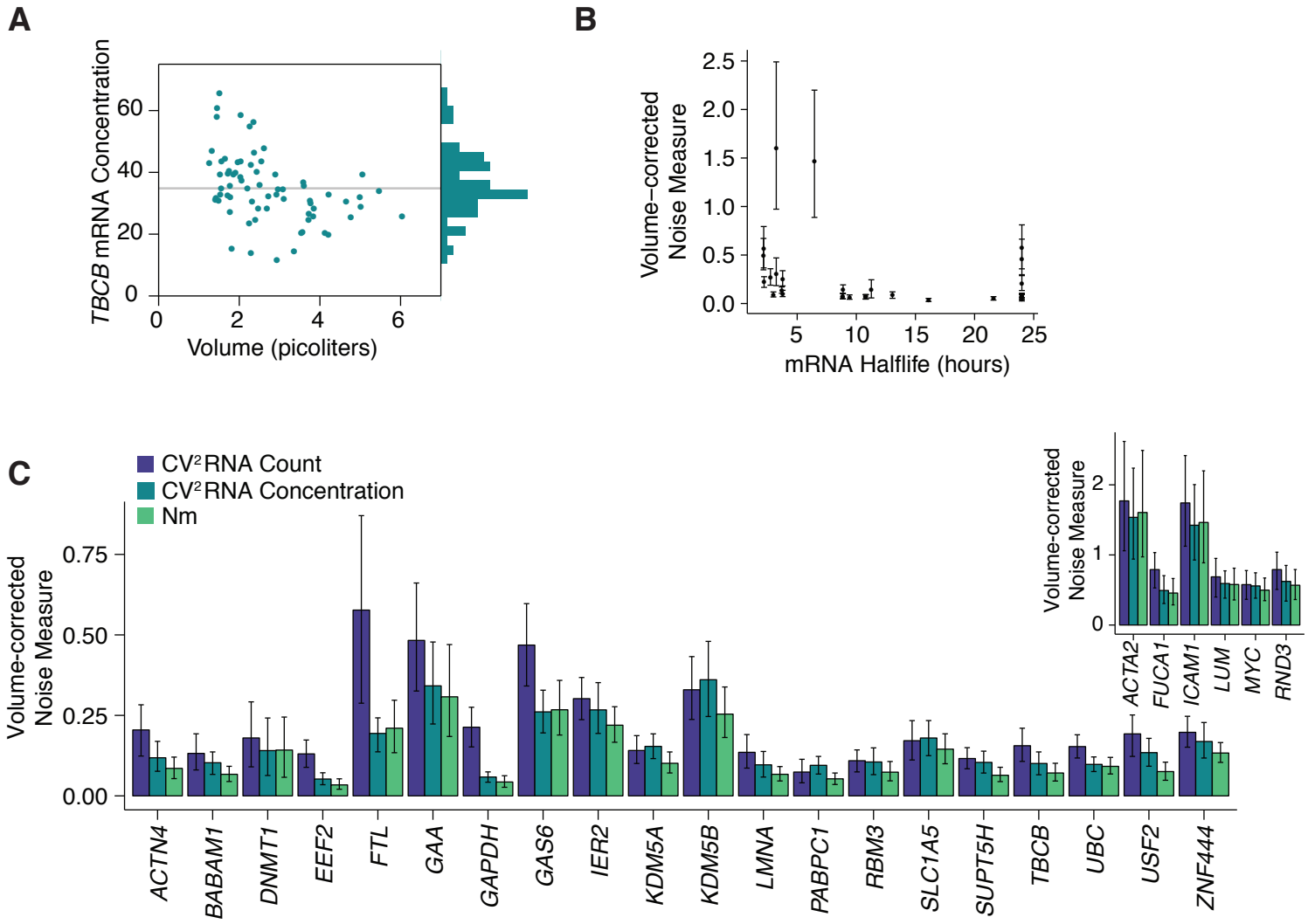

Figure S5, related to Figure 7.
A. *TBCB* mRNA concentration vs. volume. These data are the same as in (Fig. 6A), but each is normalized by volume. Histogram indicates distribution of mRNA concentration. Gray line indicates average concentration. Data are from a combination of two biological replicates.
B. We compared volume-corrected noise measure and mRNA half-life. We obtained half-life values from Tani et al., Genome Res. (2012). We find that volume-corrected noise measure does not depend strongly on half-life. Each data point represents one gene. For each gene, we have at least two biological replicates with at least 30 cells per replicate. Error bars represent 95% confidence intervals, calculated by bootstrapping.
C. mRNA CV, concentration CV, and volume-adjusted CV for cycling primary fibroblast cells. Inset shows genes that exhibit higher cell-to-cell variability in RNA, and had values too high for main axes. Generally, mRNA CV is highest, followed by concentration CV and volume-adjusted CV. Error bars represent 95% confidence intervals by bootstrapping.

Supplemental Figure 6

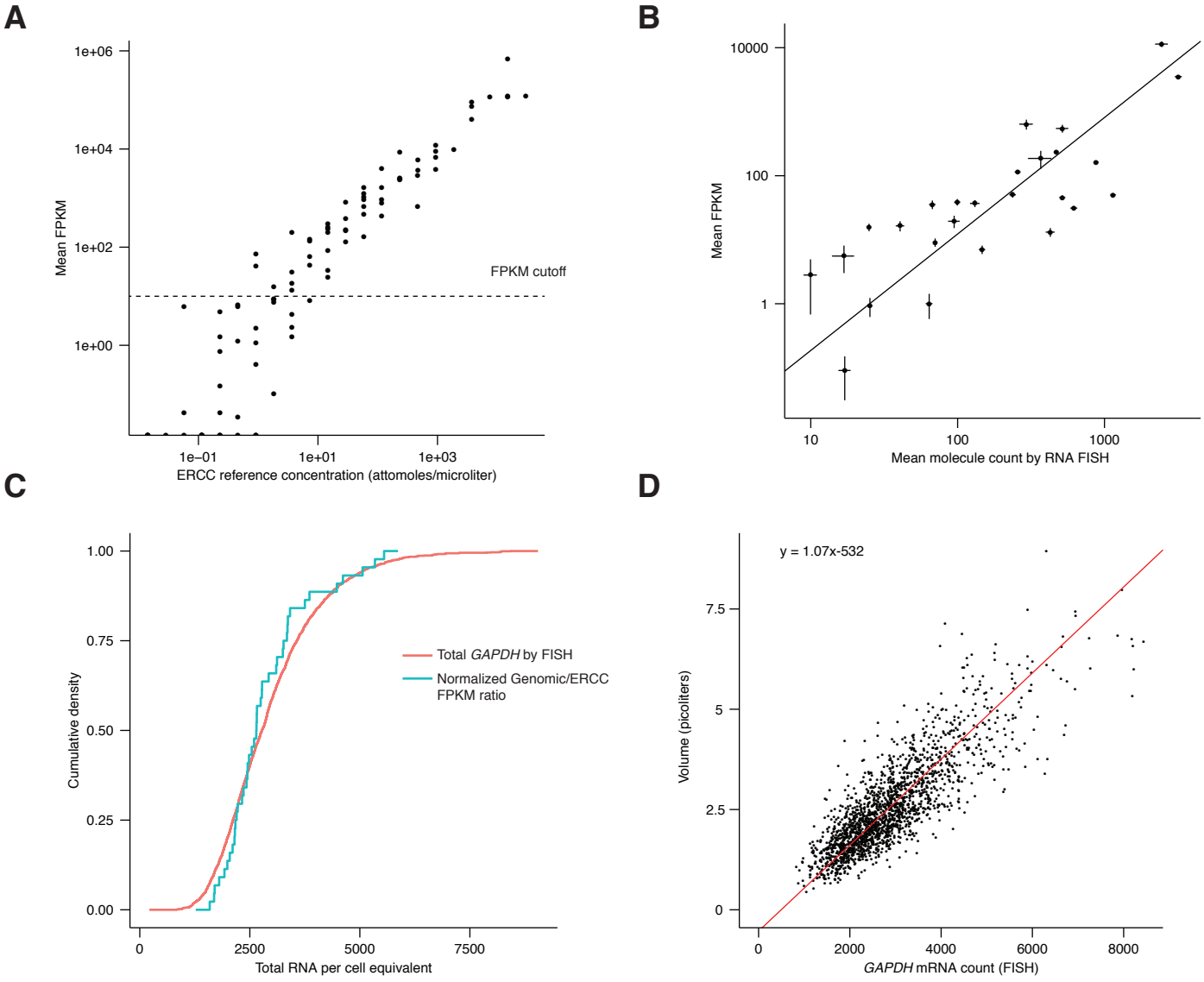

Figure S6, related to Figure 7.

A. Mean FPKM and known concentrations for each of the ERCC reference transcripts. Each point represents a single ERCC transcript and is an average over all 96 samples. Below 10 FPKM, we begin to see significant dropouts, and therefore choose and FPKM of 10 as our cutoff for “reliable” measurements. All other data in the manuscript is taken from genes having greater than 10 FPKM.

B. Mean count as measured by RNA FISH vs. mean FPKM from single-cell RNA sequencing. Each point represents a single gene and is an average over 44 single cells for single-cell sequencing, and an average over at least two biological replicates with at least 30 cells apiece for RNA FISH. These data suggest that an FPKM of 1 corresponds to approximately 23.2 transcripts per cell, as measured by RNA FISH in our cells, although the relationship between RNA FISH counts and FPKM scales nonlinearly (FPKM ~ (FISH)1.7, see Methods). We used this fitted relationship between RNA FISH count and FPKM to transform FPKM into transcript counts. Error bars represent SEM.

C. Comparison of “total RNA” distributions from single-cell sequencing and RNA FISH. Data represent a collection of total RNA measurements from single cells. We assume that total GAPDH mRNA counts by RNA FISH are proportional to total RNA. For sequencing data, we use the ratio of reads mapped to genomic loci to reads mapped to ERCCs as a proxy for total RNA. We scaled this ratio to have the same mean as the distribution of total RNA by RNA FISH. After scaling, the distributions are similar, suggesting that our method for measuring total RNA via the ratio of genomic to ERCC transcripts is sound.

D. Mapping between total RNA count (here, total GAPDH mRNA in single cells) and volume, as measured by RNA FISH. Each point represents a single cell. We use this mapping to convert total RNA from sequencing experiments to actual volume. The red line is the best fit, as computed by principle components analysis.

Supplemental Figure 7

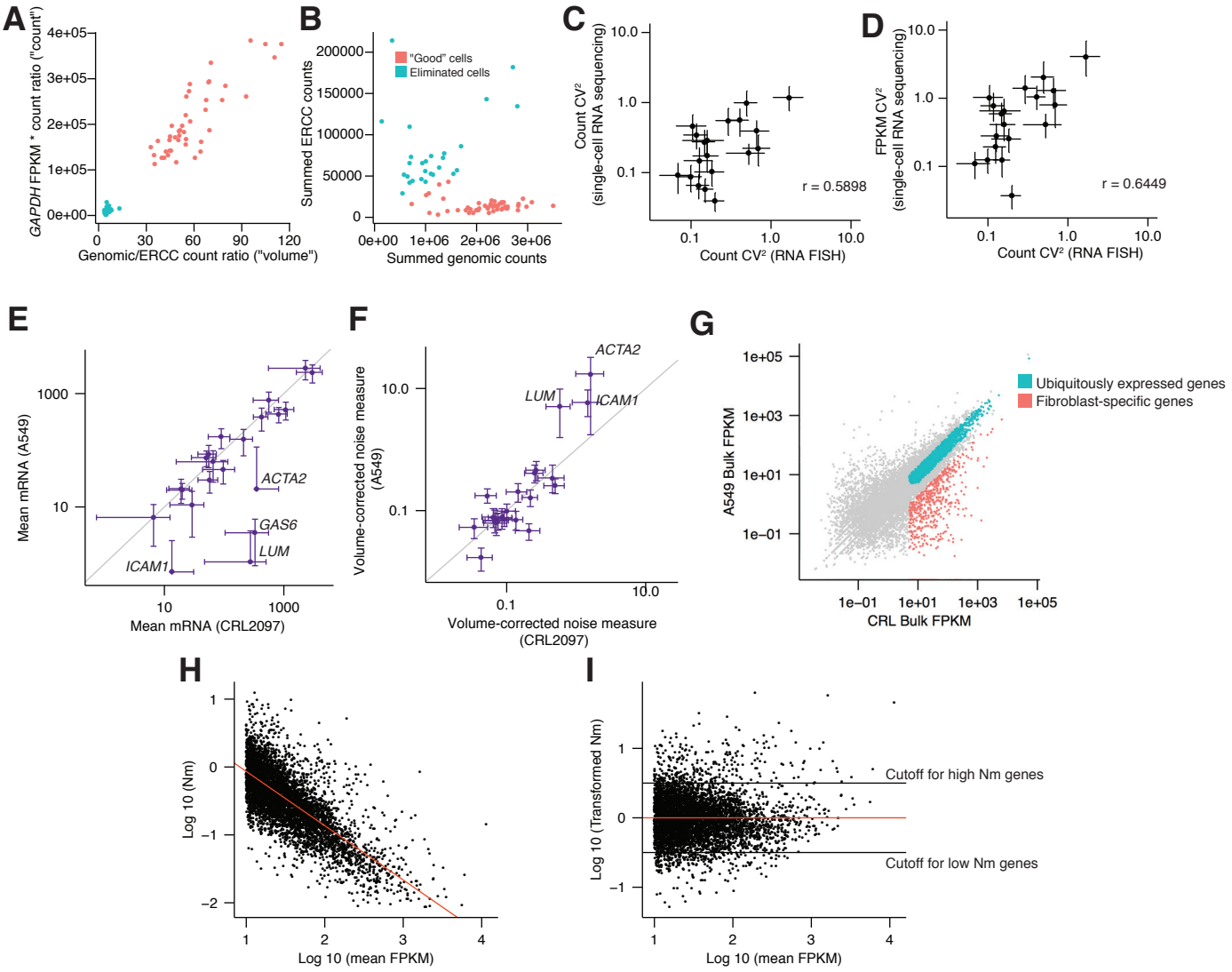

Figure S7, related to Figure 7.

A. "Count" vs. "volume" for *GAPDH* from single-cell sequencing data. We define "volume" as the ratio between genomic reads and ERCC reads for each cell. This quantity is more representative of total RNA, which we know to be roughly proportional to volume, although the relationship is not exactly proportional due to volume-independent transcription (see main text Figure 1). We observed two clearly distinct classes of cells, those with a volume range that matches what we see by imaging and RNA FISH and those that have very low volumes. For unknown reasons, these cells ended up with a considerably higher ratio of ERCC reads than genomic reads, and we eliminated them from our subsequent analyses.

B. ERCC counts vs. genomic counts for the cells that we kept and those we eliminated.

C. Correspondence between CV^2 of RNA FISH counts and CV^2 of inferred counts from single-cell sequencing (transforming FPKM as in Supplemental Figure 6 and Methods). Each data point represents a single gene.

D. Same as C, except using CV^2 of FPKM from single-cell sequencing instead. Correlation is slightly higher than in C.

For C and D, error bars represent 95% confidence intervals, calculated by bootstrapping.

E. Average mRNA counts in cycling primary fibroblasts and A549 cells, calculated using RNA FISH. Gray line indicates a 1:1 correspondence. Error bars represent standard error of the mean.

F. Volume-corrected noise measure in cycling primary fibroblast and A549 cells, calculated using RNA FISH. Gray line indicates a 1:1 correspondence. Nm calculated by bootstrapping; error bars represent 95% confidence interval. Data in E and F for each gene is a combination of at least two biological replicates, with at least 30 cells per replicate.

G. FPKM measurements from bulk RNA-sequencing in primary fibroblast and A549 cells. Each point represents one gene. We classified genes as "ubiquitously expressed" if they had >5 FPKM in both cell types and differed by less than a factor of 2 in FPKM across the two cell types. We considered genes "fibroblast specific" if they had >5 FPKM in fibroblasts and their FPKM was greater than five times higher in fibroblasts that A549 cells.

H. Single-cell RNA-sequencing data in primary fibroblast cells. Each point represents one gene. We used the method described in Supplemental Figure 6 and Methods to calculate Nm for each gene. We observe that higher abundance genes typically have lower Nm values. Red line indicates best fit line.

I. The same data as in H, but transformed to remove the volume-dependence from Nm. Red line here is the transformed fit line from H. We use this transformed data to select volume-matched "low Nm" and "high Nm" genes using a cutoff of Nm=0.5 and Nm=-0.5, respectively. We selected 307 high Nm genes and 257 low Nm genes. Note that these high Nm genes actually have a higher mean abundance (FPKM=196.5) than the low Nm genes (FPKM=55.4), thus showing that the observed differences in noise levels are not due to the overall increase in noise in genes of low abundance.

Table S1: List of all oligonucleotides used for single molecule RNA FISH. Each oligonucleotide is a DNA oligonucleotide with sequence given 5' to 3'. The sequence name gives the target of the oligonucleotide probe set.

Supplemental Experimental Procedures

Cell culture:
We grew primary human foreskin fibroblast cells (CCD-1079Sk, ATCC CRL-2097™) and A549 cells (human lung carcinoma, A549, ATCC CCL-185™) in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and 50U/mL penicillin and streptomycin (Pen/Strep). To create quiescent cells, we grew primary fibroblast cells in DMEM with Pen/Strep, without FBS for seven days. We cultured WM983b-GFP-NLS cells (WM983b is a human melanoma from the lab of Meenhard Herlyn) in Tu2% media (78% MCDB media, 20% Leibovitz's L-15 media, 2% FBS, and 1.68 mM CaCl2). The WM983b-GFP-NLS contains EGFP fused to a NLS driven by a cytomegalovirus promoter that we stably transfected into the parental cell line. Before imaging, we plated cells on two-well Lab-Tek chambered coverglasses.

RNA fluorescence *in situ* hybridization and imaging:
We performed single molecule RNA FISH on the samples as described previously (Femino et al., 1998; Raj and Tyagi, 2010; Raj et al., 2008). Briefly, we fixed the cells in formaldehyde or methanol, performed RNA FISH using the specified pools of oligonucleotides, then washed and stained nuclei with DAPI. We fixed most cells in this study using formaldehyde, but used methanol for the experiments involving transcription site quantification because it resulted in more accurate transcription site detection. We stained the actin cytoskeleton with Phalloidin-Alexa 488 (Life Technologies) to detect cell boundaries.

We typically co-stained with sets of probes targeting many different mRNA. Typically, we used exon probes labeled in Alexa 594, introns with Cy3, Cyclin A2 with Atto 647N (which labels cells in S, G2 and M phase (Eward et al., 2004)) and *GAPDH* with Atto 700. To distinguish cells in S phase from G2 (Robertson et al., 2000; Whitfield et al., 2002), we labeled Histone H4 mRNA with Atto 647N and Cyclin A2 mRNA with Atto 700 (see Supplemental Fig. 3). Supplemental data file 1 lists all sequences of oligonucleotide probes.

We imaged the cells with a Nikon Ti-E equipped with appropriate filter sets. We took a series of optical z-sections, each 0.2-0.35 microns high, that spanned the vertical extent of the cell.

Image analysis and quantification:
We manually identified cell boundaries and counted and localized RNA spots using custom software written in MATLAB (Raj and Tyagi, 2010; Raj et al., 2008). We estimate the technical error in our RNA count determination to be at most 15%.

To compute the volume of a cell, we detected the 3D positions of a highly abundant mRNA by RNA FISH. We selected only the points that defined the outer boundary of the cell by examining each point and its neighbors within a 4μm radius. We kept only the points that had a higher z-position than their neighbors (signifying the top of the cell) or points that had no neighbors within 180 degrees (signifying the side of the cell). Once we had the points, we interpolated the points to identify a smooth representation of the cell surface. We repeated this in both an upward and

downward direction to identify the top and bottom of the cell. We calculated the volume of the cell by summing the heights between the top and bottom.

Calculating the volume in this manner will always result in an underestimation of the actual volume. To correct for this bias, we first computed the outline of the cell as described above. We then dilated this hull, filled it with the same number of randomly distributed points, and then repeated the algorithm on this new set of points. If this volume matched that computed with the actual spots in the cell, we then computed the volume by integrating between the top and bottom boundaries of the dilated hull.

We used *GAPDH* mRNA as the primary mRNA for our volume determinations, but the volume computation did not depend on the number of spots identified nor on the choice of volume-filling gene (Supplemental Fig. 1). We limited ourselves to the cytoplasmic volume by removing a vertical cylinder corresponding to the nuclear outline. This procedure does exclude the cytoplasmic volume above and below the nucleus, but that region only comprised a very small proportion of the total cytoplasmic volume.

We identified transcription sites through intron/exon probe colocalization. We manually annotated transcription sites by visually inspecting images of intron and exon probes to determine instances of colocalized signal. To determine spot intensity, we identified the z-plane of maximum intensity in a 0.375μm-square region around the manually selected spot. We defined the intensity as the difference between this maximum value and a background value. For the background value, we used the median intensity in a 3.75μm-square annular region around the maximum intensity point. Note that transcription site intensity need not necessarily linearly relate to transcriptional burst size (Senecal et al., 2014).

RNA degradation:
We measured RNA degradation by inhibiting transcription for four hours by applying actinomycin D at 1ug/ml. We measured degradation of *UBC* and *IER2* mRNA because they exhibited a strong correlation with volume while having a half-life short enough to enable us to observe substantial degradation within four hours of actinomycin D treatment while avoiding non-specific effects at longer times.

We used a model to determine whether degradation was volume-dependent (degradation ~ 1/V) or volume-independent (degradation ~ constant). We first fit the untreated mRNA vs. volume data with a line having zero intercept. If degradation is volume-independent, we expect the treated cells to also be well-fit by a line having zero intercept, where the slope is determined by the untreated fit and an exponential decay term:

$$m_{4h}(V) = s_0 V e^{-\gamma t},$$

where m_{4h} is the mRNA count after 4 hours of treatment, s_0 is the slope of the untreated data ($t = 0$), γ is the decay constant (degradation rate), and t is the treatment time. Note that here γ is the only fit parameter and is independent of volume.

If degradation is volume-dependent, the equation becomes:

$$m_{4h}(V) = s_0 V e^{-\gamma t/V}.$$

Here, γ/V is the decay constant (degradation rate), but γ itself is independent of volume and is the fit parameter.

The line and curve described by these equations are the fits to the raw data, and the decay constants γ and γ/V are the fits we show to the calculated decay constants that we show in Fig. 3A-B.

We calculated the actual decay constant (Fig. 3A-B) for each cell measured at the 4 hour timepoint assuming exponential decay:

$$m_{4h}(V) = m_{0h}(V)e^{-\gamma t},$$

where we approximate $m_{0h} = s_0 V$, and γ could in principle be either volume-dependent or -independent.

***LMNA* siRNA knockdown:**

We used an siRNA targeting *LMNA* (Cat. #: AM16708, ID: 40502) at 30nM and a "scramble" control siRNA (Cat. #: AM4611) at 30nM. We incubated primary fibroblast cells with the siRNA for 72 hours. We verified protein knockdown via Western blot, using the SC-20680 (rabbit) antibody and a goat-anti-rabbit 680 RD secondary (Licor 926-68071).

Heterokaryon formation:

We created heterokaryons by separately culturing primary fibroblast cells and WM983b-GFP-NLS cells. Once the plates were 70-90% confluent, we trypsinized the cells, resuspended them in DMEM Complete media, and combined half of each plate of cells in a 15ml tube. We pelleted the cells and resuspended in PEG for 2 minutes. We added media over the course of five minutes to allow cells to fuse, then plated the cells onto two-well chambered coverglasses (Lab-Tek) and fixed the cells after 12 hours.

We identified heterokaryons as cells with two nuclei that expressed both GFP (WM983b-GFP-NLS only) and *GAS6* mRNA (primary fibroblast) by RNA FISH. We eliminated all homokaryons from our analyses.

Fractionation and RNA polymerase II Western blot:

We performed cell fractionation as described in (Bhatt et al., 2012) and based on (Wuarin and Schibler, 1994) with modifications. We conducted all subsequent steps on ice or at 4°C and in the presence of 25 μM α-amanitin (Sigma, A2263) and Protease inhibitors cOmplete (Roche, 11873580001) according to manufacturer's instructions. We pre-chilled all buffers on ice before use. We grew primary fibroblast cells to a confluency of 90%. We removed media and washed plates twice with 1x PBS before scraping cells into 1x PBS. We collected cells by centrifuging at 500 g for 10 min. We gently resuspended the cell pellet corresponding to $1x10^7$ cells in 200 μl cytoplasmic lysis buffer (0.15% NP-40, 10 mM Tris-HCl pH 7.0, 150 mM NaCl). We incubated the cell lysate for 5 min on ice, layered it onto 500 μl sucrose buffer (10 mM Tris-HCl pH 7.0, 150 mM NaCl, 25% sucrose) and centrifuged at 16,000 g for 10 min. We carefully removed the

supernatant (600 µl) corresponding to the cytoplasmic fraction. We gently resuspended the nuclei pellet in 400 µl nuclei wash buffer (0.1% Triton-X-100, 1 mM EDTA, in 1x PBS) and centrifuged it at 1,500 g for 1 min. We removed the supernatant and gently resuspended the pellet in 200 µl glycerol buffer (20 mM Tris-HCl pH 8.0, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT). Next, we added 200 µl nuclei lysis buffer (1% NP-40, 20 mM Hepes pH 7.5, 300 mM NaCl, 1M Urea, 0.2 mM EDTA, 1 mM DTT), vortexed, incubated on ice for 2 min and centrifuged at 18,500 g for 2 min. We carefully removed the supernatant corresponding to the nucleoplasmic fraction (350 µl) and added 250 µl 1x PBS/Protease inhibitors cOmplete to adjust the volume for Western blot experiments (described below). We resuspended the chromatin pellet in 600 µl chromatin resuspension solution (25 µM α-amanitin, Protease inhibitors cOmplete, in 1x PBS).

We monitored the success of cell fractionation by Western blot analyses. For Western blot analyses, we probed membranes with the following primary antibodies: Pol II (F-12, Santa Cruz Biotechnology; directed against the N-terminal region of Rpb1), Pol II Ser2-P (3E10, Active Motif), Pol II Ser5-P (3E8, Active Motif), Histone 2B (FL-126, Santa Cruz Biotechnology), U1 snRNP70 (C-18, Santa Cruz Biotechnology) and GAPDH (6C5, Applied Biosystems). Next, we probed membranes with Cy5- and Alexa Fluor 647-conjugated secondary antibodies (Cy5 goat anti-mouse, A10524; Cy5 goat anti-rabbit, A10523; Cy5 goat anti-rat, A10525; Alexa Fluor 647 rabbit anti-goat, A21446; Life Technologies), and scanned using a Typhoon 9400 scanner (GE Healthcare). We quantified fluorescent signals with ImageJ 1.47v software.

Triptolide:
We degraded RNA polymerase II in primary fibroblast cells by incubating cells in 100nM triptolide for one hour, then fixed cells in methanol (control cells remained untreated).

Cell size verification:
To check that fixation did not alter cell size, we monitored the size of cells through the fixation and permeabilization process by fixing cells while on the microscope stage. We monitored cell area by taking images in brightfield, and we monitored cell height by coating the cells with fluorescent beads and imaging them in a series of optical *z*-sections. We took images of the same cells after 10 minutes of fixing in 4% formaldehyde and after 30 minutes of permeabilization in ethanol. We calculated cell area by segmenting the cells as usual, and we determined height by identifying the plane of the bottom of the cell and the plane of the top of the cell (the last plane where beads remain motionless) and subtracting the two values.

Quantification of cell-to-cell variability:
We developed a phenomenological metric for cell-to-cell variability that takes into account both volume-correlated and volume-independent contributions to mRNA numbers per cell (see supplemental note for derivation and further information). We also used a model of transcriptional bursting with volume-dependent transcription that enabled us to quantify transcriptional parameters from population distributions of mRNA counts and volumes.

Repli-seq analysis:

We accessed Repli-seq data from Hansen et al. 2010 (Hansen et al., 2010) using the UW Repli-seq track on the UCSC Genome Browser.

Bulk RNA Sequencing:
We sequenced total RNA from primary fibroblast cells. We used the NEB Next Ultimate Library Preparation Kit for Illumina and the Ribo-Zero Magnetic Gold Kit. We used 50b single-end reads and sequenced each of two replicates at a depth of 10-15M reads. We aligned reads to hg19 using STAR's included annotation. We quantified reads per gene using HTSeq and a RefSeq hg19 annotation. We calculated FPKM for each gene using R. All sequencing data is available at GEO accession number GSE66053.

Single-cell RNA Sequencing:
We isolated 96 single cells, lysed, and performed first- and second-strand synthesis on a Fluidigm C1 Single-Cell Auto Prep System using a large size chip. We spiked in ERCC (External RNA Controls Consortium) RNA controls, Mix 1 (Ambion 4456740) at a concentration of 1:10,000 before adding the cells to the C1. We prepared cDNA libraries using the Nextera XT DNA Sample Preparation Kit (Illumina, PN FC-131-1096) and used 96 paired barcodes from the Nextera XT DNA Sample Preparation Index Kit (96 Indices, 385 Samples) (Illumina, PN FC-131-1002) following the abbreviated Fluidigm protocol for the Nextera XT kit. We sequenced the samples on a NextSeq 500 using 75b paired-end reads to a depth of ~1-2M reads per sample. To quantify sequencing data, we aligned reads to hg19 (using STAR's included annotation) and the ERCC reference transcripts. We quantified reads per gene using HTSeq and a RefSeq hg19 annotation. All sequencing data is available at GEO accession number GSE66053.

Single-cell RNA Sequencing Calibration and Analysis:
We independently calculated ERCC and genomic FPKM for each sample, normalizing to the total number of reads mapped to ERCC loci or genomic loci, respectively. All FPKM data shown for endogenous genes is this genomic FPKM. For each cell, we considered the ratio of total genomic reads to total ERCC reads to be proportional to the total starting amount of mRNA in that cell.

We sequenced 96 wells total, of which 5 were "control" wells that contained no cells and 14 were wells containing fixed cells. We excluded these 19 cells from the analysis. Further, we excluded 12 cells that had fewer than 1 million total reads, and 21 cells that had a genomic/ERCC read ratio of less than 30. We performed all further analyses on the 44 remaining cells.

Transform read ratio to volume: We assumed that the ratio of genomic/ERCC reads for each sample was proportional to the total mRNA in each cell. We also assumed that, for our RNA FISH measurements, total *GAPDH* mRNA counts were proportional to the total amount of mRNA in each cell. The distributions for total mRNA obtained in this manner were similar between RNA FISH and single-cell RNA sequencing, but had different means. We therefore normalized the sequencing data to have the same mean as the RNA FISH distribution. From our RNA FISH dataset, we have many co-measurements of *GAPDH* mRNA (total mRNA) and

volume from which we establish a transformation equation between total mRNA and volume. We obtained this transformation equation using PCA, or orthogonal regression. Using this equation, we transformed total mRNA obtained through sequencing into actual volume in picoliters.

Transform FPKM to molecule count: FPKM is more a measure of mRNA concentration than mRNA count, as it is normalized to total reads. To get a measure more similar to mRNA count, for each cell, we multiplied each gene's FPKM by the genomic/ERCC count ratio ("volume") of the cell. For each gene in our RNA FISH dataset, we fit the log of the seq "counts" and the log of the actual counts from RNA FISH by orthogonal regression. We then used this transform to convert the FPKM of all genes to their RNA FISH count equivalent. Note that, because we fit in log space, the transform between FPKM and count is nonlinear, and actually scales as approximately FPKM ~ (RNA FISH)$^{1.7}$.

Once we had our single-cell sequencing data in terms of RNA FISH count and volume in picoliters, we calculated Nm as described for RNA FISH. We performed all of our sequencing analysis in R.

***C. elegans* growth and imaging:**

We grew N2 (wild type) and CB502 (*sma-2* mutant) *C. elegans* on NGM agar plates with OP50 lawns, kept at 20º C. Every 2-3 days, we transferred a small portion of each strain to new plates to prevent overgrowth.

We released the worms off of the plates using phosphate buffered saline (PBS) solution, then fixed with 4% formaldehyde for 45 minutes. We permeabilized and stored the worms in 70% ethanol. We performed the RNA FISH protocol, then mounted the sample between a slide and coverglass before imaging.

We manually identified head and gonad boundaries and counted and localized RNA spots using custom software written in MATLAB.

To compute the volume of each worm segment, we multiplied the area of the segment by the height of the segment (thus approximating the segment as a prism). We determined the height by taking the vertical difference between the highest and lowest RNA spots' positions, as determined by our software. We determined the number of cells in each segment through manually counting the DAPI-stained nuclei.

We obtained data from multiple segments. When combining the data (number of mRNA spots per volume or per nucleus), we weighted each segment by its volume.

Model of diffusible *trans* factor for volume:DNA ratio sensing

Here, we outline a fairly generic model for how a diffusible *trans* factor may transmit information on the ratio of volume to DNA to lead to increased transcription in larger cells irrespective of DNA content. The primary assumptions are that the factor is predominantly localized to the nucleus, the factor is required for mRNA transcription, and the cellular concentration of the factor is roughly constant irrespective of cellular volume (i.e., the total amount of factor is proportional to cellular volume). RNA polymerase II holoenzyme satisfies these conditions, although we do not claim that RNA polymerase II is the factor.

The model assumes binding of the factor to the DNA, and that only bound factor can result in productive transcription. The goal of the model is to provide a basis for the empirical finding that larger cells have increased transcription from the same absolute number of DNA molecules. Our model encompasses two broad categories of mechanism that would lead to a perfectly linear scaling of transcription with cytoplasmic volume: (1) The factor is sequestered entirely in the nucleus, and so if the nucleus doesn't change with cellular volume, the concentration of the factor in the nucleus will be proportional to the total amount of factor. Thus, the factor will be proportionally more bound to the DNA in a larger cell than a smaller cell, producing more transcription. (2) The factor is a purely "limiting" factor in the sense that it has a very high affinity for DNA and the number of binding sites exceeds the amount of factor. In this situation, essentially all available factor will be bound to the DNA, and so for each gene, there would be proportionally more transcription in larger cells because more factor would be bound to DNA. These mechanisms are not necessarily mutually exclusive. The model incorporates affinity and nuclear volume as parameters, and so encompasses both of these potential mechanisms.

Briefly, the conclusion we derive from our model is that both scenarios pose viable mechanisms for scaling transcription with cellular volume. That said, we overall mildly favor scenario 1. Our data show that nuclear volume increases somewhat with nuclear size, which the model predicts should lead to a slight decrease in transcription in larger cells, and thus a higher concentration of mRNA in smaller cells, which is precisely what we observe. Moreover, there is a rough quantitative agreement between the degree of increased nuclear volume and the higher concentration of mRNA in smaller cells. Definitively proving that the cell follows scenario 1 of our model will require further experiments.

We begin with a few definitions. We use quantities within brackets to denote concentration (molecules per volume) and quantities without brackets to denote number of molecules per cell. For instance, $\text{factor}_{\text{free}}$ is the number of free molecules of the factor, while $[\text{factor}_{\text{free}}]$ is the number of free molecules per unit volume. $\text{factor}_{\text{DNA}}$ denotes the number of factor molecules instantaneously bound to DNA, $\text{factor}_{\text{total}}$ is the total amount of factor in the cell/nucleus, and DNA is the number of binding sites on the DNA for the factor in the nucleus. K_{DNA} is the binding affinity of the factor for a particular gene. The cellular volume is given by V, and the nuclear volume by V_{nuclear}. Thus, given our assumption of proportionality, we define p_{factor} to be a constant such that $\text{factor}_{\text{total}} = p_{\text{factor}} V$.

The total factor is given by

$$\text{factor}_{\text{total}} = \text{factor}_{\text{free}} + \text{factor}_{\text{DNA}}, \tag{1}$$

Dividing by the nuclear volume, we arrive at a relationship between concentrations:

$$[\text{factor}_{\text{total}}] = [\text{factor}_{\text{free}}] + [\text{factor}_{\text{DNA}}]\,. \tag{2}$$

The binding affinity is defined via mass action as

$$K_{\mathrm{DNA}} = \frac{[\mathrm{factor}_{\mathrm{free}}]\,[\mathrm{DNA}]}{[\mathrm{factor}_{\mathrm{DNA}}]}. \tag{3}$$

and may be different for different genes owing to promoters having different numbers of binding sites for the factor or different binding affinities.

Thus, the total concentration of the machinery bound to DNA is

$$[\mathrm{factor}_{\mathrm{DNA}}] = \frac{([\mathrm{factor}_{\mathrm{DNA}}] - [\mathrm{factor}_{\mathrm{total}}])\,[\mathrm{DNA}]}{K_{\mathrm{DNA}}}. \tag{4}$$

Solving for $[\mathrm{factor}_{\mathrm{DNA}}]$, we find:

$$[\mathrm{factor}_{\mathrm{DNA}}] = \frac{[\mathrm{factor}_{\mathrm{total}}]\,[\mathrm{DNA}]}{K_{\mathrm{DNA}} + [\mathrm{DNA}]}. \tag{5}$$

In the limiting case where $K_{\mathrm{DNA}} = 0$, we expect all of the factor to be bound to DNA, and in that case, we find $[\mathrm{factor}_{\mathrm{DNA}}] = [\mathrm{factor}_{\mathrm{total}}]$, as expected.

Relating concentrations to volumes yields

$$[\mathrm{factor}_{\mathrm{total}}] = \frac{p_{\mathrm{factor}} V}{V_{\mathrm{nucleus}}}, \tag{6}$$

where p_{factor} is the proportionality constant defined earlier. Similarly,

$$[\mathrm{DNA}] = \frac{\mathrm{DNA}}{V_{\mathrm{nucleus}}}. \tag{7}$$

Hence,

$$[\mathrm{factor}_{\mathrm{DNA}}] = \frac{p_{\mathrm{factor}} \frac{V}{V_{\mathrm{nucleus}}} \frac{\mathrm{DNA}}{V_{\mathrm{nucleus}}}}{K_{\mathrm{DNA}} + \frac{\mathrm{DNA}}{V_{\mathrm{nucleus}}}}. \tag{8}$$

Simplifying,

$$[\mathrm{factor}_{\mathrm{DNA}}] = \left(\frac{1}{V_{\mathrm{nucleus}}}\right) \frac{p_{\mathrm{factor}} \cdot V \cdot \mathrm{DNA}}{K_{\mathrm{DNA}} \cdot V_{\mathrm{nucleus}} + \mathrm{DNA}}. \tag{9}$$

Because $[\mathrm{factor}_{\mathrm{DNA}}] = \mathrm{factor}_{\mathrm{DNA}}/V_{\mathrm{nucleus}}$, we can solve for the total amount of transcriptional machinery bound to DNA:

$$\mathrm{factor}_{\mathrm{DNA}} = \frac{p_{\mathrm{factor}} \cdot V \cdot \mathrm{DNA}}{K_{\mathrm{DNA}} \cdot V_{\mathrm{nucleus}} + \mathrm{DNA}}. \tag{10}$$

In the limiting case $K_{\mathrm{DNA}} = 0$ here, we find that $\mathrm{factor}_{\mathrm{DNA}}$ is directly proportional to volume, and equal to the total amount of factor in the nucleus irrespective of nuclear volume. However, in the case where K_{DNA} is not zero, then the volume of the nucleus will result in deviations from perfect scaling of transcription with cellular volume. Intuitively, if the volume of the nucleus increases somewhat in larger cells, then the concentration of the factor and the DNA will decrease and hence the amount of factor bound to DNA will be somewhat less than it would be otherwise. In that case, larger cells would have somewhat less transcription than would be expected in the case of perfect scaling of transcription with cellular volume, which fits with our experimentally observed

volume-independent transcript abundance (i.e., decreased mRNA concentration in larger cells). We also observed that nuclear volume is somewhat greater in larger cells. Thus, it was possible, at least qualitatively, that the increase in nuclear volume could explain the apparent decrease in mRNA concentration in larger cells. We thus wanted to check whether there is a quantitative agreement between our observed relationship between nuclear volume and cytoplasmic volume and the increased mRNA concentration in smaller cells, which would establish the plausibility of such a model.

As mentioned, our measurements show that nuclear area and cellular volume positively correlate. Approximating nuclear volume by raising nuclear area to the 3/2 power, we find a linear relationship between nuclear "volume" and cellular volume ($V_{\text{nucleus}} \propto a + bV$), with y-intercept $a = 2169$ femtoliters (95% C.I. = (1923, 2381)), and slope $b = 0.9354$ (95% C.I. = (0.8313, 1.042)). It is important to note that the while the relationship is well-fit by a line, the line does not pass through zero, and so nuclear volume is *not* directly proportional to total cellular volume. Using this linear relationship, we can express the total amount of factor bound to DNA as a function of cellular volume:

$$\text{factor}_{\text{DNA}}(V) = \frac{p_{\text{factor}} \cdot V}{\left(\tilde{a} + \tilde{b}V\right) + 1}, \tag{11}$$

where $\tilde{a} = \frac{K_{\text{DNA}}}{\text{DNA}} \cdot a$ and $\tilde{b} = \frac{K_{\text{DNA}}}{\text{DNA}} \cdot b$. The ratio a/b $(= \tilde{a}/\tilde{b})$ has units of volume, and is geometrically equivalent to the x-intercept of the line of best fit between nuclear volume and cellular volume.

We now wanted to check whether the volume-independent transcription we observed in our mRNA-volume plots would quantitatively agree with this model. Because the factor is required for transcription and only transcribes when bound to DNA, then each gene essentially grabs a fixed proportion of the amount of factor bound to DNA. (This fraction will depend on the specific regulation of the gene.) So the total transcription of a gene will be proportional to $\text{factor}_{\text{DNA}}$. Thus, lumping together this proportionality constant along with mRNA production and degradation and other associated constants into a constant c, the relationship between RNA and volume is given by:

$$\text{RNA}(V) = c \cdot \text{factor}_{\text{DNA}}(V). \tag{12}$$

We should be able to fit our RNA vs. volume data using the above equation to obtain estimates for $\tilde{a}$ and $\tilde{b}$, in particular their ratio, which is directly comparable to the ratio a/b.

We did so for three genes and found fitting parameters:

	UBC	$ZNF444$	$EEF2$
$\tilde{a}$ (fL)	1.578	95.68	0.2747
95% C.I. ($\tilde{a}$) (fL)	(0.8524, 2.461)	(70.23, 127.9)	(-0.1518, 0.5630)
$\tilde{b}$	1.936×10^{-4}	0.01495	0.0003658
95% C.I. ($\tilde{b}$)	(-7.617×10^{-5}, 4.595×10^{-4})	(0.004163, 0.02394)	(0.0002680, 0.0005879)
$\tilde{a}/\tilde{b}$ (fL)	7044	6446	744.4
95% C.I. ($\tilde{a}/\tilde{b}$) (fL)	(-62743, 111200)	(2988, 28110)	(-253.5, 2064)

For the fit of nuclear area to volume, we find the ratio $a/b = 2329$ femtoliters, with a 95% confidence interval of (1844, 2851). We note that the ratios $\tilde{a}/\tilde{b}$ for all of our genes are of that same order of magnitude, albeit with large error. This result suggests that the above equation for $\text{factor}_{\text{DNA}}(V)$ may be the equation governing the production of mRNA in cells. This model

provides an explanation that is quantitatively consistent with our data for why smaller cells exhibit proportionally slightly more transcription than larger cells—nuclei in small cells are slightly smaller than those in large cells, increasing the concentration of $\text{factor}_{\text{DNA}}(V)$, and therefore increasing transcription.

Our results are consistent with RNA polymerase II holoenzyme being the factor. RNA polymerase II is required for transcription, transcribes when bound to DNA, and is almost exclusively localized to the nucleus. Also, most reports indicate that most RNA polymerase II in the nucleus is not specifically bound to DNA. Based on that fact, one would expect that increased nuclear size should lead to slight under-transcription, as we observed. Our analysis shows that this relationship is quantitatively plausible. Further studies will be required to rigorously establish that RNA polymerase II holoenzyme is the factor that connects volume/DNA ratio to transcription.

Note, however, that in many situations such as in early embryogenesis, nuclear size does change dramatically as a function of cellular volume. In these situations, the mechanism we describe would face a challenge because the concentration of RNA polymerase II holoenzyme in the cell's nucleus would remain the same after division (and the associated decrease in cellular volume), leading to over-transcription. The limiting factor model (scenario 2, with K_{DNA} very small) would not suffer from these issues. It is possible that some intermediate scenario is at play in early embryogenesis.

Another problem with these models is the potential for runaway positive feedback, in which a random increase in the factor would lead to more production of the factor, thus leading to runaway transcription. For this reason, we expect that the cell maintains strong control of factor levels to avoid these issues. Ultimately, a complete understanding of factor dynamics will likely require adding growth to models of transcriptional homeostasis.

Computing volume-corrected noise measure from single-cell mRNA and volume measurements

We define *volume-corrected noise measure* as the cell-to-cell expression variability in mRNA levels that cannot be accounted for by cell-to-cell differences in volume. Throughout the section, we denote the variance of a random variable X by σ_X^2.

Let m and V be random variables denoting single-cell mRNA level and volume, respectively. The expected number of mRNA transcripts in a cell given its volume V is assumed to increase linearly with V, i.e.,

$$\langle m|V\rangle = a + bV \implies \langle m\rangle = a + b\langle V\rangle, \tag{13}$$

where $\langle . \rangle$ represents the expected value, and a, b are gene-specific constants (volume-independent and volume-correlated transcript abundance, respectively). From (13), the covariance between m and V is given by

$$Cov(m, V) = \langle mV\rangle - \langle m\rangle\langle V\rangle = \langle (a + bV)V\rangle - (a + b\langle V\rangle)\langle V\rangle = b\sigma_V^2. \tag{14}$$

The extent of cell-to-cell variability in mRNA counts that can be accounted for by volume is

$$\sigma^2_{\langle m|V\rangle} = \sigma^2_{a+bV} = b^2\sigma_V^2, \tag{15}$$

which using (14) can be written

$$\sigma^2_{\langle m|V\rangle} = bCov(m, V). \tag{16}$$

Volume-corrected noise measure Nm defined as

$$Nm := \frac{\sigma^2_m - \sigma^2_{\langle m|V\rangle}}{\langle m\rangle^2} \tag{17}$$

is obtained as follows using (16)

$$Nm = CV^2_m - \frac{bCov(m, V)}{\langle m\rangle^2} = CV^2_m - S\frac{Cov(m, V)}{\langle m\rangle\langle V\rangle}, \tag{18}$$

where CV^2_m represents the total variability in mRNA levels measured by its Coefficient of Variation (CV) squared and

$$S = \frac{b\langle V\rangle}{\langle m\rangle} = \frac{b\langle V\rangle}{a + b\langle V\rangle}. \tag{19}$$

Noise measure in a two-state promoter model

Consider two alleles, where each allele transitions independently between active and inactive states with rates k_{on} and k_{off}. We assume that the transcription rate from the active state increases linearly with cell volume V. We first compute CV^2_m (mRNA coefficient of variation squared) for the case where transcription is independent of volume, and then extend it to the volume dependent case.

Transcription rate independent of volume

Let the transcription rate from active state be k_m. Then, the steady-state first and second-order moment of the mRNA level m is given by

$$\langle m\rangle = \frac{2G_{on}k_m}{\gamma_m}, \quad \langle m^2\rangle = \langle m\rangle + \frac{\gamma_m(1 - G_{on})\langle m\rangle^2}{2(G_{on}\gamma_m + k_{on})} + \langle m\rangle^2, \tag{20}$$

where

$$G_{on} = \frac{k_{on}}{k_{on} + k_{off}} \tag{21}$$

is the fraction of time an allele is in the active state, and γ_m is the mRNA degradation rate. Note that the factor two in (20) arises due to the presence of two alleles. This results in

$$CV^2_m := \frac{\langle m^2\rangle - \langle m\rangle^2}{\langle m\rangle^2} = \frac{1}{\langle m\rangle} + \frac{\gamma_m(1 - G_{on})}{2(G_{on}\gamma_m + k_{on})}. \tag{22}$$

Transcription rate dependent on volume

We assume $k_m = a + bV$, where volume V is a random variable with mean $\langle V \rangle$ and variance σ_V^2. Based on (20),

$$\langle m|V\rangle = \frac{2G_{on}(a+bV)}{\gamma_m}, \quad \langle m^2|V\rangle = \langle m|V\rangle + \frac{\gamma_m(1-G_{on})\langle m|V\rangle^2}{2(G_{on}\gamma_m + k_{on})} + \langle m|V\rangle^2. \tag{23}$$

Unconditioning on the volume we obtain

$$\langle m\rangle = \frac{2G_{on}(a+b\langle V\rangle)}{\gamma_m} \tag{24a}$$

$$\langle m^2\rangle = \langle m\rangle + \frac{\gamma_m(1-G_{on})\left\langle\langle m|V\rangle^2\right\rangle}{2(G_{on}\gamma_m + k_{on})} + \left\langle\langle m|V\rangle^2\right\rangle. \tag{24b}$$

Using (23)

$$\left\langle\langle m|V\rangle^2\right\rangle = \left\langle\left(\frac{2G_{on}(a+bV)}{\gamma_m}\right)^2\right\rangle = \langle m\rangle^2(1+S^2CV_V^2), \tag{25}$$

where the mean mRNA count $\langle m\rangle$ is given by (24a), S is given by (19) and CV_V^2 is the volume CV^2. Substituting (25) in (24b)

$$\langle m^2\rangle = \frac{\gamma_m(1-G_{on})\langle m\rangle^2(1+S^2CV_V^2)}{2(G_{on}\gamma_m + k_{on})} + \langle m\rangle^2(1+S^2CV_V^2) + \langle m\rangle. \tag{26}$$

Above equation yields

$$CV_m^2 := \frac{\langle m^2\rangle - \langle m\rangle^2}{\langle m\rangle^2} = \frac{1}{\langle m\rangle} + \frac{\gamma_m(1-G_{on})(1+S^2CV_V^2)}{2(G_{on}\gamma_m + k_{on})} + S^2CV_V^2. \tag{27}$$

As expected, (27) reduces to (22) when $CV_V^2 = 0$. In (27), the first term represent Poissonian noise in mRNA level due to random-birth death of individual mRNA molecules. The second term is the noise contribution from stochastic promoter switching. Variation in mRNA levels due to cell-to-cell differences in cell volume is represented by the last term. Removing the last term, we obtain the noise measure as

$$Nm = \frac{1}{\langle m\rangle} + \frac{\gamma_m(1-G_{on})(1+S^2CV_V^2)}{2(G_{on}\gamma_m + k_{on})}. \tag{28}$$

Estimating promoter transition rates between active and inactive states

Noise measures obtained from single-cell mRNA count and volume measurements are used to estimate promoter transition rates k_{on} and k_{off} using (28). To correct for measurement noise, we take into account a 15% error in mRNA counting. From (21) and (28)

$$G_{on} = \frac{k_{on}}{k_{on} + k_{off}} \tag{29a}$$

$$Nm = \frac{1}{\langle m\rangle} + \frac{\gamma_m(1-G_{on})(1+S^2CV_V^2)}{2(G_{on}\gamma_m + k_{on})} + CV_{count}^2, \tag{29b}$$

Supplementary Table I: Average promoter dwell-time (30) obtained from the noise measure (Eq. (29)) or the total mRNA expression variability (Eq. (31)). mRNA half-lives and dwell times are reported in hours.

Gene	G_{on}	CV_m^2	Nm/CV_m^2	mRNA half-life	T_{on}, T_{off} from Nm	T_{on}, T_{off} from CV_m^2
ACTN4	0.65	0.14	0.4	13	6.1, 3.3	40.5, 21.8
GAPDH	0.65	0.23	0.17	24	4.9, 2.6	331, 178
EEF2	0.75	0.18	0.17	16	3.2, 1.1	1443.6, 481.2
FTL	0.3	0.58	0.32	24	6.3, 14.6	45.4, 105.9
ICAM1	0.05	0.8	0.8	6.5	0.5, 9.7	0.8, 15.4
ACTA2	0.2	1.18	0.9	3	5.1, 20.3	7.9, 31.6
LUM	0.15	0.7	0.75	24	7.8, 44.3	12.7, 72.5
SUPTH5	0.3	0.12	0.58	10	0.45, 1.1	1.4, 3.3

where $CV_{count}^2 = 0.15^2 = 0.0225$ represents the mRNA counting error. In the above equations, quantities Nm, S (defined in (19)), CV_V^2 (volume CV^2), $\langle m \rangle$, G_{on} are computed from data for a given gene. Using mRNA half-life information from literature, rates k_{on} and k_{off} can be estimated by solving (29). The average promoter dwell time in the active and inactive state is given by

$$T_{on} = \frac{1}{k_{off}}, \quad T_{off} = \frac{1}{k_{on}}, \tag{30}$$

respectively, and reported in Table I under the column "T_{on}, T_{off} from Nm". Since there may be other unaccounted sources of noise in gene expression, these dwell time estimates should be considered an upper bound on their actual values.

We contrast the above estimates to the scenario where all the mRNA expression variability is assumed to arises from transcriptional bursting. From (22), k_{on} and k_{off} in that case would be estimated by solving

$$G_{on} = \frac{k_{on}}{k_{on} + k_{off}} \tag{31a}$$

$$CV_m^2 = \frac{1}{\langle m \rangle} + \frac{\gamma_m(1 - G_{on})}{2(G_{on}\gamma_m + k_{on})} + CV_{count}^2. \tag{31b}$$

Since $CV_m^2 > Nm$, dwell times obtained from (31) (see "T_{on}, T_{off} from CV_m^2" in Table I) are significantly larger than those obtained from (29). For example, using the GAPDH noise measure, we estimate $T_{on} = 4.9$ hours. However, if one ignores the contribution of cell volume in driving intercellular variation in GAPDH mRNA, the mean dwell time in the active state is obtained to be $13 - 14$ days (331 hours) from (31).